

182066

**STIC-Biotech/ChemLib**

---

**From:** Chernyshev, Olga  
**Sent:** Monday, March 13, 2006 4:38 PM  
**To:** STIC-Biotech/ChemLib  
**Subject:** 10/736,936, sequence search request

Please search SEQ ID NO: 2 in regular and pending databases.  
Thank you very much!

*Olga N. Chernyshev, Ph.D.*  
*AU 1649*  
*REM 3C89*  
*2-0870*  
*mail 4C70*

\*\*\*\*\*

Searcher: \_\_\_\_\_  
Searcher Phone: \_\_\_\_\_  
Date Searcher Picked up: \_\_\_\_\_  
Date completed: 3-14-06  
Searcher Prep Time: \_\_\_\_\_  
Online Time: \_\_\_\_\_

\*\*\*\*\*

Type of Search  
NA# \_\_\_\_\_ AA# \_\_\_\_\_  
S/L: \_\_\_\_\_ Oligomer: \_\_\_\_\_  
Encode/Transl: \_\_\_\_\_  
Structure #: \_\_\_\_\_ Text: \_\_\_\_\_  
Inventor: \_\_\_\_\_ Litigation: \_\_\_\_\_

\*\*\*\*\*

Vendors and cost where applicable  
STN: \_\_\_\_\_  
DIALOG: \_\_\_\_\_  
QUESTEL/ORBIT: \_\_\_\_\_  
LEXIS/NEXIS: \_\_\_\_\_  
SEQUENCE SYSTEM: \_\_\_\_\_  
WWW/Internet: \_\_\_\_\_  
Other (Specify): \_\_\_\_\_

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: March 14, 2006, 10:48:24 ; Search time 44 Seconds  
(without alignments)  
1511.041 Million cell updates/sec

Title: US-10-736-936-2

Perfect score: 3598  
Sequence: 1 MGNQHLNKTAEKQPSBENCK.....SLNKKHFPVPSAGADSETHC 691

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Length	ID	Description	
1	2351	65.3	689	2	JC7286	liver-specific org
2	1411.5	39.2	670	2	JC7616	organic anion tran
3	1396.5	38.8	670	2	A49580	mediates transport
4	1334.5	37.1	670	2	JC7581	organic anion-tran
5	940.5	26.1	643	2	A41120	prostaglandin tran
6	894	24.8	682	2	JC7385	multispecific orga
7	723	20.1	809	2	T16448	hypothetical prote
8	686.5	19.1	1451	2	T16388	hypothetical prote
9	630	17.5	758	2	T26595	hypothetical prote
10	618.5	17.2	690	2	T27357	hypothetical prote
11	609	16.9	674	2	T21217	hypothetical prote
12	527	14.6	482	2	T17250	hypothetical prote
13	495.5	13.8	655	2	T16538	hypothetical prote
14	178	4.9	431	2	B81254	probable transmem
15	178	4.9	507	2	B84616	hypothetical prote
16	173.5	4.8	403	2	B84063	multidrug resistan
17	171.5	4.8	746	2	T05899	hypothetical prote
18	170.5	4.7	469	2	H97064	probable sugar-pro
19	160	4.4	487	2	B97884	probable tartrate
20	159	4.4	410	2	B84998	hypothetical prote
21	158	4.4	606	2	T27072	hypothetical prote
22	157.5	4.4	423	2	S74046	probable sugar tra
23	156	4.3	461	2	H64636	proline/betaine tr
24	153.5	4.3	425	2	AC2909	MFS permease (limo
25	149	4.1	613	2	SS7687	probable membrane
26	144	4.0	436	2	A81443	probable transmem
27	143.5	4.0	452	2	A71877	proline/betaine tr
28	143	4.0	471	2	H90502	phosphate transpor
29	142	3.9	457	2	D71144	hypothetical prote

30	138.5	3.8	541	2	T40518	major facilitator
31	138.5	3.8	567	2	A71619	membrane transport
32	137.5	3.8	422	2	H69839	multidrug resistan
33	137.5	3.8	613	2	T27077	hypothetical prote
34	137	3.8	433	2	B84566	hypothetical prote
35	137	3.8	513	2	T12997	hypothetical prote
36	135.5	3.8	474	2	B83719	multidrug resistan
37	134.5	3.7	457	2	H85059	probable sugar tra
38	134.5	3.7	464	2	F69587	L-arabinose transp
39	134.5	3.7	522	2	T41604	probable membrane
40	134	3.7	391	2	D71676	bicyclomycin resis
41	134	3.7	419	2	B90446	permease (imported
42	133.5	3.7	489	2	D85434	hypothetical prote
43	133	3.7	387	2	C86640	multidrug efflux t
44	133	3.7	429	1	B41502	arsenical pump mem
45	133	3.7	497	2	S66834	probable membrane

ALIGNMENTS

RESULT 1  
JC7286  
liver-specific organic anion transporter-1 - mouse  
C:Species: Mus musculus (house mouse)  
C:Date: 18-Aug-2000 #sequence\_revision 18-Aug-2000 #text\_change 09-Jul-2004  
C:Accession: JC7286  
R:Ogura, K.; Choudhuri, S.; Klaassen, C.D.  
Biochem. Biophys. Res. Commun. 272, 563-570, 2000  
A:Title: Full-length cDNA cloning and genomic organization of the mouse liver-specific c  
A:Reference number: JC7286  
A:Accession: JC7286  
A:Molecule type: mRNA  
A:Residues: 1-689 <OGU>  
A:Cross-references: UNIPROT:Q9JUL3; UNIPARC:UPI0000023434; DDBJ:AB031959  
C:Genetics:  
A:Gene: lat-1  
C:Keywords: glycolysis; transmembrane protein

Query Match 65.3%; Score 2351; DB 2; Length 689;  
Best Local Similarity 63.8%; Pred. No. 4,8e-159;  
Matches 442; Conservative 97; Mismatches 144; Indels 10; Gaps 6;

QY	1	MGNQHLNKTAEKQPSBENCKRYCNGLKMFLEALSFSFAKTGLGATIMKSIHRRRE	60
DB	1	MGNQHLNKTAEKQPSBENCKRYCNGLKMFLEALSFSFAKTGLGATIMKSIHRRRE	58
QY	61	ISSSLVGFIDGFEIGNLIVIVFVSFKLHRLPKLIGICFIMGIGVLTALPHFPMGY	120
DB	59	ISSSLVGFIDGFEIGNLIVIVFVSFKLHRLPKLIGICFIMGIGVLTALPHFPMGY	118
QY	121	YYSKRTNIDSSSENSTSTLCINQIISLNRASPEIVKGLKESGSMYIVFVGNL	180
DB	119	YYSKRTNIDSSSENSTSTLCINQIISLNRASPEIVKGLKESGSMYIVFVGNL	175
QY	181	RIGETPIVPLGYSYDDFAKSGHSLYGLINAIAMGPIIGFTLGSFSMYVDIGVY	240
DB	176	RIGETPIVPLGYSYDDFAKSGHSLYGLINAIAMGPIIGFTLGSFSMYVDIGVY	235
QY	241	DLSIRITPTDSRWGAMWMLFVSGLFSSISPEFLPQTPNPKQKRKASLSHLVLE	300
DB	236	DLSIRITPTDSRWGAMWMLFVSGLFSSISPEFLPQTPNPKQKRKASLSHLVLE	295
QY	301	TNDEKQDTANLNQGNITKAVTGFQSKSLITNPPLVMPFLLTLQVSSYIGAFYVF	360
DB	296	TNDEKQDTANLNQGNITKAVTGFQSKSLITNPPLVMPFLLTLQVSSYIGAFYVF	355
QY	361	KYVEOOYGGSSKANILGVTITIPASGPFAGYVIRKFKNTVGIASFCTAMSLS	420
DB	356	KYVEOOYGGSSKANILGVTITIPASGPFAGYVIRKFKNTVGIASFCTAMSLS	415
QY	421	PYLLVFFILCENKSVAGLTMTYDGNPVTSHRDVPLSYCNSDCDESGWEPVCGNGIT	480
DB	421	PYLLVFFILCENKSVAGLTMTYDGNPVTSHRDVPLSYCNSDCDESGWEPVCGNGIT	480

Db 416 FLYSTFLICENKAPGLTLITDGNAPVDSHIDPLSTYCNSDCTCDKQMEPVCGENGYT 475  
 QY 481 YISPLCAGCKSSGNKK--PIYFYNCSCEBVTGLQNRNYSANHLGECPRDDACTRKYFPV 538  
 Db 476 YISPLCAGCKSPRGDKLNMIBFYDCSCVSGSGFGKGNHARLGECPR-DCKCTKYFYFI 534  
 QY 539 ALOVALNFPSSALGCTSHWMLIVKIYVPELKSALAGHSHVIRLQGLIAPYFGALIDPT 598  
 Db 535 TFGVILISFTALGSLTLMILIRSVQPELKSGLMGFHSLVKTLGGIILAPVYGGALIDRT 594  
 QY 599 CTKSTNNCGTRGSCRTYNSTSPSRVYLAGLSWLRVSLVLYIILYAMKKYQOEODINA 658  
 Db 595 CKNKSTYSGARGACALYNSRLFGMIYVGLSTAKTPILLIYVALIYMKRKRKRDNKI 654  
 QY 659 SENG-SYMDENALIESLNKKR-FVPSAGADSET 689  
 Db 655 LENGKFTDEGNPEPVNNNGYSCVPSEDKNSET 687

## RESULT 2

JC7616  
 organic anion transport polypeptide 2 - mouse  
 C/Species: Mus musculus (house mouse)  
 C/Date: 30-Jun-2001 #sequence\_revision 30-Jun-2001 #ext\_change 09-Jul-2004  
 C/Accession: J07616  
 R/Ogura, K.; Choudhuri, S.; Klaassen, C.D.  
 Biochem. Biophys. Res. Commun. 281, 431-439, 2001  
 A/Title: Genomic organization and tissue-specific expression of splice variants of mouse  
 A/Reference number: J07616; M0ID:21092843; PMID:11181066  
 A/Accession: J07616  
 A/Molecule type: mRNA  
 A/Residues: 1-670 <OGU>  
 A/Cross-references: UNIPROT:Q9B9P6; UNIPARC:UPI00002711E; DDBJ:AB031814  
 C/Comment: This protein with twelve transmembrane domains, glycosylation sites and prote  
 C/Genetics:  
 A:Gene: oalp2  
 C/Keywords: transmembrane protein  
 F/21-39/Domain: transmembrane #status predicted <TM1>  
 F/59-80/Domain: transmembrane #status predicted <TM2>  
 F/87-105/Domain: transmembrane #status predicted <TM3>  
 F/157-179/Domain: transmembrane #status predicted <TM4>  
 F/192-214/Domain: transmembrane #status predicted <TM5>  
 F/244-266/Domain: transmembrane #status predicted <TM6>  
 F/317-339/Domain: transmembrane #status predicted <TM7>  
 F/354-376/Domain: transmembrane #status predicted <TM8>  
 F/387-409/Domain: transmembrane #status predicted <TM9>  
 F/514-536/Domain: transmembrane #status predicted <TM10>  
 F/550-571/Domain: transmembrane #status predicted <TM11>  
 F/600-622/Domain: transmembrane #status predicted <TM12>

Query Match 39.2%; Score 1411.5; DB 2; Length 670;  
 Best Local Similarity 43.2%; Pred. No. 2.2e-92;  
 Matches 290; Conservative 134; Mismatches 214; Indels 33; Gaps 12;

QY 27 LKMPFLAALISPAKTGAIIMKSIHIERREISSLVGFDGSEIGNLVIYFVY 86  
 Db 19 IKAFLALICAYVSKLSGTYMMSMLTQIRQGIPTSVGLNGSEIGNLVIYFVY 78  
 QY 87 FSGKLHAPKLIIGICFIMGIGVLTALPHFPMGYRYSKETNIDSENSTSTLTCLNQ 146  
 Db 79 FGTGLHAPKLIIGICFIMGIGVLTALPHFPMGYRYSKETNIDSENSTSTLTCLNQ 136  
 QY 147 ILSINRASPRIYKGLCKESGYSYMIYVFMGNMLRGIGETPIYVPLGSLYIDDPKXGHS 206  
 Db 137 TQTL---KPTQDPTECVCKEMSLMWIYLVGNIRKNGEPIPLGSIYIEDPAKSENBP 193  
 QY 207 LYGILNALAMIPITIGFTLGSFSKMYVDIGYVDSTIRITPTDSRWGAWMLNLVSG 266  
 Db 194 LVIGILFTGWTIGFTLGLLSSCANIYVDTSVNTDILITPTDRWGAWMIGFLVCA 253  
 QY 267 LFSIISIPFPFLPQTPNKRQKERKASLSLHVLSTNDEKQDTANLTNQGKNITKXVTGP 326  
 Db 254 GVNIILTSIPFPFPKTLKEGLQDNGDGT---ENAKKEKHREKIKENRGLTKD---FF 306

QY 327 QSEFKLTNPILYMFVTLTLQVSSYICAFYTVFKVVEQOQPSKANIILGVTITPFE 386  
 Db 307 LPMKSLISCPRIYMIFFLISVIOVNAFINSTPMPTKLEQOYKSTBIYFLMGLYMLPEI 366  
 QY 387 ASGMFLGYIIRKPKNTVGIKFSCTAVMSLPIYLF---FILCENKSVAGLMTYD 443  
 Db 367 CIGYILGILMKKFKI-TYKKAAYIGFW--LSLTETILSFSVYIMTCNDPFPAGLITTSIE 423  
 QY 444 G-NNPVTSHRDVPLSYCNSDCNDSQWEPVCGNNGITYISPLCAGCKSSGNKKPIYFV 502  
 Db 424 GYQHPLIYVENNV-LADCNFKSGCLTNMTDPVCGDNGLSYMSACLAGECKSVGTGNTMVPQ 482  
 QY 503 NCSGLEVTGLQNRNYSANHLGECPRDDACTRKYFPFPAILOVALNFPSSALGCTSHWMLIVKI 562  
 Db 483 NCSGLQSSG---NMSAVGLCDKGECCKNKLQYFLIISITICFIFSLGALIGYVLLKC 538  
 QY 563 VQPELKSALGFSHVIKALGILAPYFGALIDPTCIKSTNNCGTRGSCRTYNSTSPS 622  
 Db 539 MKSEKSLGVGLHTPCMKRLIGGIPAPYFGALIDRTCLHMTLKGEPGACMIDINSFR 598  
 QY 623 RYVLGLSMLRVSSLVLYIILYAMKKYQOEODINASENGVMDAANL-----ESLNKK 677  
 Db 599 RYVLGHPALRGASFLPALFILMRKFQFPEDIDSSDT---DPAEMKLTAKESKCTIV 654  
 QY 678 HVPVPSAGADSE 688  
 Db 655 HNSPTQNDGE 665

## RESULT 3

A49580  
 mediates transport of organic anion - rat  
 C/Species: Rattus norvegicus (Norway rat)  
 C/Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #ext\_change 09-Jul-2004  
 C/Accession: A49580  
 R/Jacquemin, E.; Hagenbuch, B.; Stieger, B.; Wolkoef, A.W.; Meier, P.J.  
 Proc. Natl. Acad. Sci. U.S.A. 91, 133-137, 1994  
 A/Title: Expression cloning of a rat liver Na(+)-independent organic anion transporter.  
 A/Reference number: A49580; M0ID:94105118; PMID:8278353  
 A/Accession: A49580  
 A/Status: preliminary; translated from GB/EMBL/DDBJ  
 A/Molecule type: mRNA  
 A/Residues: 1-670 <RES>  
 A/Cross-references: UNIPROT:P46720; UNIPARC:UPI0000135446; GB:U19031; NID:9410310; PIDN:  
 C/Genetics:  
 A:Gene: oalp

Query Match 38.8%; Score 1396.5; DB 2; Length 670;  
 Best Local Similarity 42.9%; Pred. No. 2.5e-91;  
 Matches 289; Conservative 125; Mismatches 231; Indels 29; Gaps 11;

QY 22 RYCNGLKMPFLAALISPAKTGAIIMKSIHIERREISSLVGFDGSEIGNLVIY 81  
 Db 14 RPSKKAFLVLTSLTCACTKSLGYSVMNMLTQIRQFISTSVGLNGSEIGNLVIY 73  
 QY 82 VFPVSYGSKLHAPKLIIGICFIMGIGVLTALPHFPMGYRYSKETNIDSENSTSTLT 141  
 Db 74 VFPVSYGSKLHAPKLIIGICFIMGIGVLTALPHFPMGYRYSKETNIDSENSTSTLT 131  
 QY 142 CLINQILSNRASPRIYKGLCKESGYSYMIYVFMGNMLRGIGETPIYVPLGSLYIDDPK 201  
 Db 132 CMENRQTL---KPTQDPABCVKEMSLMWICVMGNIRIGIGETPIYVPLGSLYIEDPAK 188  
 QY 202 EGHSSLYILNALAMIPITIGFTLGSFSKMYVDIGYVDSTIRITPTDSRWGAWMLN 261  
 Db 189 SENSPLIYILKNGKAGGIFGLLGSYCAQIYVDIGSVNTDILITPTDRWGAWMIG 248  
 QY 262 PLVSGFLSIISIPFPFLPQTPNKRQKERKASLSLHVLSTNDEKQDT---ANLTNQGKNI 318  
 Db 249 FLVACAVNIILTSIPFPFLKALPKKGQDS-----NVAVTIDGKVEKKGGARBNIGI 301  
 QY 319 TKXVTFQSFQSKILTNPILYMFVTLTLQVSSYICAFYTVFKVVEQOQPSKANIIL 378

```

Db      302 TND---FLTPMKRLPCNPITMLFILTSLVQVNGFIKFTFLPKRTLEQOIGKSTAEKIFILI 358
Qy      379 GTITIPFASGMLGGYIIKKFKLNTVGLAKESCPTAVMSLSFYLLYFPILCENKSVAGL 438
Db      359 GYVSLPEPICGLYGLIGFIMKKFKITVKAAYLAFLGLSVPEYLLFLCHFLMLTCDMAVAGL 418
Qy      439 TMTYDGNMNTVSHRDVPLSYCNSDCNCSQWEPVCGNNGITYIISPCLAGCKSSGKKKPY 498
Db      419 TTSYGVQHQHVLVESKVLADCNTRCSCTNTWDPVCGDNGVAYMSACLGAQCKKPFVGTGN 478
Qy      499 IYFVNCSCLEVTGLQNRNYSALHGCPRDADCTRKPFYFVAIOVNLFPFSLAGSTHWL 558
Db      479 WYFQDCSCISLG---NSSAVLGLCKKPEPCANRLOVFLITIIISFYLSLTALPGYNV 534
Qy      559 IYKIVQPELKLALGFHSMVIRALGGILAPIYFGLIDTTCIKWSTNCGTRGSCRTYNS 618
Db      535 PLRCVSEKESLGVGLHTFCIRVPAGIPAPVYFGALIDRTCLHMGTCLKCGQACGMVDI 594
Qy      619 TSPSRVYLGSLMRLVSS-LVLYIIITYAMKKKYQ---EKDINASENGSYMDEANLESIN 674
Db      595 NSFRHLYLGLPIALRGSSYLPAFFILI--LMRKFPQPGDIDSSAYDHTMM-LGEKESSE 651
Qy      675 KKHGFVPSAGADSE 688
Db      652 TDVHGSPOVENDGE 665

```

## RESULT 4

```

JCT581
organic anion-transporting polypeptide-5 - mouse
C/Species: Mus musculus (house mouse)
C/Date: 30-Jun-2001 #sequence_revision 30-Jun-2001 #text_change 09-Jul-2004
C/Accession: JCT581
R/Choudhuri, S.; Ogura, K.; Klaassen, C.D.
Biochem. Biophys. Res. Commun. 280, 92-98, 2001
A/Title: Cloning, expression, and ontogeny of mouse organic anion-transporting polypeptide
A/Reference number: JCT581; PMID: 21092571; PMID:11162483
A/Contents: Kidney
A/Accession: JCT581
A/Molecule type: mRNA
A/Residues: 1-670 <CHO>
A/Cross-references: UNIPROT:Q99494; UNIPARC:UPI00000279D2; GB:AF203701; GB:AF213260
C/Comment: This protein, an ortholog of rat, is a kidney-specific organic anion transporter.
A/Keywords: kidney; transmembrane protein; transport protein
A/Gene: oacp-5
C/Keywords: kidney; transmembrane protein; transport protein

```

```

Query Match      37.1%; Score 1334.5; DB 2; Length 670;
Best Local Similarity 42.2%; Pred. No. 6,3e-87;
Matches 278; Conservative 115; Mismatches 238; Indels 27; Gaps 10;

Qy      27 LKMPFLAALSPIAKTGAIMKSSIHTERRFEISSSLVGFIDGFEIGNLLVIVFVY 86
Db      19 IKVFLALIMWYISKILSGVYSTMLTQERQFNISTISVGLINGSFEKGNLLVIVFVY 78
Qy      87 FGSKLHRLPLIGICGFMIGVLTALPHFPMGYRYRSKENTINDSSNSTSLTCLINQ 146
Db      79 FGTKLHRLPLMIGVGAVMGLGCFIISLPHFLMGREY--ETIISPTSMNSNSFLCVERK 136
Qy      147 ILSLRASPEIYVKGKCLKESGSYMMIYVFMGNMLGIGETPIVPLAGLSIYIDPFAKESHS 206
Db      137 SGTL---KPTQDPACVCEIKSLMIYLVGNIGIGETPIIMPAGISTIEPFAKESNP 193
Qy      207 LYLGLILNAMIIGPIITGLTGLSLFSKMYVDIGYDLSTIRITPTDSRWGVAWMLNPLVSG 266
Db      194 LYIGILGVKMGIMPIILGYLMGPFCANITYVDTSVNTDILTPTDTRWGAWMIGFLVCA 253
Qy      267 LFSIISSIPREFLPQTPPKPKQERKASLSLHLEFNDEKQANLITNCKNITTKAVTGF 326
Db      254 GNVAVTSLSPFFFPPTLPEEGIQDNGDGTEAKNK--EKIRDVAKENQD--TIKK---FF 306
Qy      327 QSPKSLITNPLVWVLLTLLQVSSYIGAFYVFKYVEQVQGPSSKANILLGITIPIF 386

```

```

Db      307 LMKMLPCNPITMLCVLTSVLQVNGVANIYKKPKLHHFGISTKAAVFLGLVTPSV 366
Qy      387 ASGMFLGGYIIKKFR--LNTVGIKAFSCPTAVMSLSFYLLYFPILCENKSVAGLMTYDG 444
Db      367 SAGVILISGRIMKMLKITLKKAINALCLFMSCLLS--LCNFMILTDDTPIAGLTTSYSG 424
Qy      445 NNPTVSHRDVPLSYCNSDCNCSQWEPVCGNNGITYIISPCLAGCKSSGKKKPYIYNC 504
Db      425 IQQSPDMERKFLSDCNTRCNCLTKTWDPVCGNNGVAYMSPCLAGCKESYGTGANMWFQNC 484
Qy      505 SCLBYTGLQNRNYSALHGCPRDADCTRKPFYFVAIOVNLFPFSLAGSTHWLKYIKYO 564
Db      485 SCIRSSG---NSSAVLGLCKKPEPCANRLOVFLITIIYFCCFFYSIATTPGMVFLRCMK 540
Qy      565 PELKSLALGFHSMVIRALGGILAPIYFGLIDTTCIKWSTNCGTRGSCRTYNSFSRV 624
Db      541 SEKSLIGLGLQAFMRFLFAGIPAPIYFGALIDRTCLHMGTCLKGEBGACRTYEVSFRL 600
Qy      625 YIGLSMRLVSSLVYIIITYAMKKKYQ-----EKDINASENGSYMDEANLESIN 676
Db      601 YLGLPALR-GSIIIPSEFPIRLIRKLOIPGDDISEIETAKPTKESSECTDMKS 657

```

## RESULT 5

```

A41120
prostaglandin transporter - rat
N/Alternate names: matrix P/G
C/Species: Rattus norvegicus (Norway rat)
C/Date: 27-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 09-Jul-2004
C/Accession: A41120
R/Hakes, D.J.; Beresney, R.
Proc. Natl. Acad. Sci. U.S.A. 88, 6186-6190, 1991
A/Title: Molecular cloning of matrix P/G: a DNA binding protein of the nuclear matrix
A/Reference number: A41120; PMID:91296785; PMID:2068100
A/Accession: A41120
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-643 <HAK>
A/Cross-references: UNIPROT:000910; UNIPARC:UPI0000135445; GB:M64862
A/Note: the authors did not translate the codons for residues 1-99 in this reference
R/Kanai, N.; Lu, R.; Satriano, J.A.; Bao, Y.; Wolkoff, A.W.; Schuster, V.L.
Science 268, 866-869, 1995
A/Title: Identification and characterization of a prostaglandin transporter.
A/Reference number: A38955; PMID:95273959; PMID:7754369
A/Contents: annotation
C/Keywords: transmembrane protein

```

```

Query Match      26.1%; Score 940.5; DB 2; Length 643;
Best Local Similarity 32.3%; Pred. No. 5,4e-59;
Matches 215; Conservative 131; Mismatches 240; Indels 79; Gaps 15;

Qy      25 NGLKMPFLAALSPIAKTGAIMKSSIHTERRFEISSSLVGFIDGFEIGNLLVIVFV 84
Db      29 SNIKVFVLCGHLGQCLLYSAVYFSSLTTEKRGSLSSSGLISLNEISNATLIPI 88
Qy      85 SYFGSKLHRLPLIGICGFMIGVLTALPHFPMGYRYRSKENTINDSSNSTSLTCL 144
Db      89 SYFGSRVVRPMIIGIGLLAAGAVLTLPPLSPPYQYTSY----DGNSSYQTDI- 142
Qy      145 NQILSNRASPFIYVKGKCLKESG-----SYMMIYVFMGNMLGIGET 186
Db      143 -----CQKHFGALPSPKCHSTVEDTHKEISSLWGLWVVAQLLAGIGIV 185
Qy      187 PIVPLGSLYIDDPFAKESHSYLGILNAMIIGPIITGLTGLSLFSKMYVDIGYDLSTIR 246
Db      186 PIQPGISYVDDFAPTSPLYISILFAIVGPAFGYLGSLVMLRIFVDYGRVDTATYN 245
Qy      247 ITPTDSRWGVAWMLNPLVSGLSFSIISSIPREFLPQTPPKPKQERKASLSLHLEFNDEK 306
Db      246 LSPGDPRIAGMWLGLLISGFLVTSLPFFFPFA-----MSRAERSV-----TAETM 296
Qy      307 QTANLTNCKNITTKAVTGFQSPKSLITNPLVWVLLTLLQVSSYIGAFYVFKYVEQ 366

```

```

Db      297 QTBEDSKRG-SLMDFIKRPFRITRLKMLNDPLFMLVVLSSQCTSSSVAGLSTPLNKELEQ 355
Qy      367 YQSPSKANILLGVITIPFASGMFLGGYIIKKPKLNTVGIKAFSCPTAVMSLFYLLVF 426
      356 YGATAVAYANFLGAVLPLPAALGMLFGGILMKRFVPLQITPRVAATITITISILCLPFL 415
Qy      427 FLLCEKSKVAGLTMVTDGANNPVTSHRDVPUSYCNSDCNDESGMPEYCGNNGITITISPLCL 486
      416 FMGCSTSAVALEV---YPPSTSSSIHQOQPPA-CRRCSCSPDSFFHVPVCGONGVEYSPCH 471
Qy      487 AGCK-----SSGKKKIVFVNCSCLEKVTGLQNNENYSAHLGECPRDDACTRKFYFAIOY 542
      472 AGCSTNTTSSSEASKETI-YLNCSC--VSG-----GSASQDRMLPHTVLRALLPST 518
Db      543 LNLFPESAL-GGTSH---VMLIKIYQPELKSLLAGFHSWIRALGILAPIYEGALIDTT 598
      519 PLISFALLIACISHNPLVMVLKVNQDEKSPALIGVQFLMLRLMLPAPSLVGLLIDSS 578
Qy      599 CLKMSNNCGTSGCTTNTSTSFRRYVLGLSSMLRVSSLVLYIIILYAMKKKQOEADINA 658
      579 CVRMWYLCGRRGACAYNDALRNRYLGLQWYVKALGTLILFPISMRMK--NREYSI 635
Qy      659 SENG 663
      636 QENT 640

```

## RESULT 6

```

JCT385
multispecific organic anion transporter, moat1 - rat
C.Species: Rattus norvegicus (Norway rat)
C.Date: 17-Nov-2000 #sequence_revision 17-Nov-2000 #text_change 09-Jul-2004
C.Accession: JCT385
C.Niethio, T.; Adachi, H.; Nakagomi, R.; Tokui, T.; Sato, E.; Tanemoto, M.; Fujiwara, K.;
Iinuma, K.; Nunoki, K.; Matsuno, S.; Abe, T.
Biochem. Biophys. Res. Commun. 275, 831-838, 2000
A.Title: Molecular identification of a rat novel organic anion transporter moat1, which
A.Reference number: JCT385
A.Contents: Brain
A.Accession: JCT385
A.Molecule type: mRNA
A.Residues: 1-682 <NIS>
A.Cross-references: UNIPROT:Q9JHT3; UNIPARC:UPI000013544B; GB:AF169409
A.Comment: This protein, a transmembrane glycoprotein, transports prostaglandin D2, leuk
rostaglandin E1, E2, thromboxane B2, and lipoic acid.
C.Genetics:
A.Gene: moat1
C.Keywords: brain; glycolysis; transmembrane protein

```

```

Query Match      24.8%; Score 894; DB 2; Length 682;
Best Local Similarity 31.5%; Pred. No. 1,2e-55;
Matches 216; Conservative 139; Mismatches 268; Indels 62; Gaps 17;

Qy      12 EAQPSNKKTRYCNGKMLFALSLSFIAKTLGAIIMKSIHIERREISSLVGFID 71
      30 DAQP-----KGMFQN-IKFFVLCHSIIQLAQMLSTGLKSIIVKRFGLSGQTSLLAA 84
Db      72 SFEIGNLVIVVSYFGSKLHPKLIIGICFTWIGIGVLTALPHFMGYRYSKETNIDS 131
      85 FNEVGNISILFVSYGRVHRPRMIGCGALIVAVAGLMLALPHFISEPRYD-HSSPR 143
Qy      132 SENSSTL-----STCLINOIISLNASPEIIVKCGCKSEGSIMMIVPMGNMLRG 183
      144 SDDFESLCLPTTMADASALSNDCSSRRETKLITWGIN-----FTAQTLGL 192
Db      184 GETPIYPLGLSYIDDPFAKSGHSLVGIINAMIPPIIGFTLGSLSFKMYVDIGVDS 243
      199 GGVPFQPFQISTYIDDPFAHNSNPLVIGILFALIMMPEGLAVGSLMLALYVDIDMPSG 252
Qy      244 TIRITPTDSRWYGMWMLNVLVSGLSFIISSIPFPLPQTPNKQKERK-----ASL 294
      253 GINTLTQDRWYGMWMLNVLVSGLSFIISSIPFPLPQTPNKQKERK-----ASL 312

```

```

Qy      295 SLHVLFTNDEKD---OTANLITNQGKNITKQNTGVFFOSFKSI-----LTNPLVYMVLLTLL 347
      313 GSKGBELSSQHPRLKQACLPOIAPDLT--VVOFKIVFPRVLRITLRHPIFLVLVLSQVC 370
Qy      348 QVSSYVAGTAYVFKVVEQYQGPSSKANILLGVITIPFASGMFLGGYIIKKPKLNTVGI 407
      371 TSSMAGTITPLFKLEKRFPSITASPNALLGCTIPLAVIGIVGVGVKXHLHSPMC 430
Qy      408 AKRSCPTAVMSLFYLYEFLICENKSVAGLTMVTDGNNPVTSHRDVPUSYCNDCNDE 467
      431 SALLCLLSLCLLSLPLEFFICGTHNIGITQDL-GAQPGRS-----LEPGCSPPCSQS 485
Qy      468 SQMPEYCGNNG-ITITISPLACGK-----SSGKKKIVFVNCSCLEKVTGLQNNENYSAHLG 522
      486 DDFNPVCDTSAVAYETTPPHACCTGRVVOALDKSQVFPYNCSC--VAG---NGTISAG 539
Db      523 ECPRDDACTRKFYFAIOVLNLFPSALGTSWMLIKIYQPELKSLLAGFHSWIRAL 582
      540 SC--ESACSRVLVPLILSLGAAVASITHTSFMLILRGVKEDEKTLAVGQFMILRLVL 597
Qy      583 GGIAPYEGALIDTTCLKMSTNNCGTSGCTTNTSTSFRRYVLGLSSMLRVSSLVLYIT 642
      598 AMWSPVHSGAIDTTCVHML-TGCRRAVCRYDHDILRNRFIQLQFFPKSGSLVCFAL 656
Qy      643 LIYAMKKKQOEADINSENGSYMDE 667
      657 VLAILRQSGREASTAYATKSSDQE 681

```

## RESULT 7

```

T16448
hypothetical protein F53B1.8 - Caenorhabditis elegans
C.Species: Caenorhabditis elegans
C.Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 09-Jul-2004
C.Accession: T16448
C.Miller, N.
submitted to the EMBL Data Library, November 1995
A.Description: The sequence of C. elegans cosmid F53B1.
A.Reference number: Z18514
A.Accession: T16448
A.Status: preliminary; translated from GB/EMBL/DBJ
A.Molecule type: DNA
A.Residues: 1-809 <ML>
A.Cross-references: UNIPROT:Q20702; UNIPARC:UPI000007D95A; EMBL:U40953; NID:G1072248; PI
A.Experimental source: strain Bristol N2; clone F53B1
C.Genetics:
A.Gene: CESP:F53B1.8
A.Map position: X
A.Introns: 39/2; 66/3; 116/3; 214/1; 321/2; 405/1; 580/2; 643/3; 750/3

```

```

Query Match      20.1%; Score 723; DB 2; Length 809;
Best Local Similarity 27.6%; Pred. No. 1.9e-43;
Matches 202; Conservative 135; Mismatches 272; Indels 122; Gaps 26;

Qy      10 TABAOPSEBK---KTRYCNGKMLFALSLSFIAKTLGAIIMKSIHIERREISS 63
      81 TESQCGICKMPKMLQCLGKQDPMVL-LCYVCISQSLVAVGVPASALSIERRFKST 139
Qy      64 SLVGFIDSGFEIGNLVIVVSYFGSKLHPKLIIGICFTWIGIGVLTALPHFMGYRY 123
      140 SHMGRIYQFDGVYLLCIPVSYFGGRHSPKPVLAAGLACMLAGSLIFAPRIMSDSY-- 197
Qy      124 SKETNIDSENSSTLSITCLINOIISLNASPEIIVKCGCKSEGSIMMIVPMGNMLRG 183
      198 -----SASQGNMSPGTCISIEHLQNDWTSPEALAQICQSNQDQSSNMLYFLFCF 250
Qy      177 GNMRLGIGTPIVPLGLSYIDDPFAKSGHSLVGIINAMIPPIIGFTLGSLSFKMYVD 236
      251 AHFLGIGTPIVPLGLSYIDDPFAKSGHSLVGIINAMIPPIIGFTLGSLSFKMYVD 310
Qy      237 IGVV---DISTRIPTDSRWYGMWMLNVLVSGLSFIISSIPFPLPQTPNKQKERK 293
      253 IGVV---DISTRIPTDSRWYGMWMLNVLVSGLSFIISSIPFPLPQTPNKQKERK 312

```

```

Db      311 FMHLPEQLKVSSEDTPTWVGAMWLSFVGVAVLPLASLP-----KVLPS 363
Qy      294 LSLHLETNDE-----KQDTANLT-----NQGKN----- 317
Db      364 LKMRHRLHDAAVANRRRTPECCMGPMGSKNTALNSDAPADNFGSKAIKKEDYTTFKM 423
Qy      318 -----ITKNVGFPGSPKS-----ILNPLVYMWVLLTLQ 348
Db      424 LNKIFEITQASLLYSIPARGGRLMYKIMLDVRHPIPLAIRILTNMLFM--IITSAMA 481
Qy      349 VSSYI--GAFYVFKVVEQOYQOPSSKANILLGVITPIFASGMFLGYIILKKFKLNTVG 406
Db      482 ISLVVTTGASSPMSKLEHGFVSVPKSNALLICVAVPMAGMCMWTGGLVNHFRINSSK 541
Qy      407 IAKESCFTAVMSLSFYLLYFFILCENKSVAGLTMTY--DGNPN--VTSHRDVL-SYC 459
Db      542 MKLFAIGLIFLSLFFSPMY-LIYCPHAPLVGVDSAYPEFDGSIIPDYNAATYSKDPBLMNSC 600
Qy      460 NSDCNDESOEWBVC-----NNGITYISPCLACKSS--GSKKPIVFNVCSCLEVTGLN 514
Db      601 NKOCTDPSBYRVCALDGDGROFTYSPCYAGCSSTYSASLKE--YTECSVP-TNTKN 657
Qy      515 RNVSAAHGCPCRDACTRKFFVFAVIAQVNLFFSALGTSHWMLIVKIYQBELKSLALGF 574
Db      658 RPRIVKKGVC--BPQTELLMFLFLFAPLSFCTFAL-AVPIISVILRTVDYNERSPALGI 714
Qy      575 HSNVIRALGGLIAPYFGLIDTTCIKWSTNNCG-TRGSCRTYNSPFSRVYLGSSMLR 633
Db      715 QWIMIVIGITPAPVLFWMFVDSCKIKYDSSCGSSGGLMYSNQYLANLFLFTISQ 774
Qy      634 VSLVLYITLI 644
Db      775 ITTLALIVL 785

```

## RESULT 8

T16388

hypothetical protein F47B1.2 - Caenorhabditis elegans

C/Species: Caenorhabditis elegans

C/Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 20-Sep-1999

C/Accession: T16388

R/Pavello, T.

submitted to the EMBL Data Library, November 1995

A/Description: The sequence of C. elegans cosmid F47B1.

A/Reference number: Z18505

A/Accession: T16388

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-1451 &lt;FAV&gt;

A/Cross-references: UNIPARC:UPI000017B9DF; EMBL:U09993; NID:G1055098; PID:G1055100; PIDN

C/Genetics:

A/Gene: CESP:F47B1.2

A/Introns: 50/1; 70/2; 108/3; 154/3; 249/3; 331/1; 522/2; 551/2; 627/2; 725/3; 773/1; 79

Query Match 19.1%; Score 686.5; DB 2; Length 1451;

Best Local Similarity 24.5%; Pred. No. 1.5e-40;

Matches 188; Conservative 134; Mismatches 267; Indels 177; Gaps 20;

```

Qy      29 MFALAALSLPIAKTGAIIKSSIIHIERRFEISSILVGFIDSGFEIGNLVIVFVSYFG 88
Db      6 IFGLGSPFVPELEAGISYINSAVQNIERQFQMSRSGFMASDFGYIPCVVFAHFG 65
Qy      89 SKLHRPKLIGIGCFIMIGVLTALPHF----- 117
Db      66 SKGNRRARWIGAGIIVISITLMLASPNFLFPGQHDINTTEVSAKKPTAGQALASNTLK 125
Qy      118 -----MGYVYSKETNID-----SSENSTGLSTCLIN 145
Db      126 QLLSYGLIDRMHMDYMLINLEDDTPIEYQPIGRPKIGQSGSSTYTIIDQMLN 185
Qy      146 QTL----- 155
Db      186 EALVALENTIANNTSGNLTSSLSLFIHRRTNTSTKDIOKIRISAAPAFPGCKLTNSLR 245

```

```

Qy      156 EIVKGCCKE--SGSYMMIYVFMGNLRIGIEPIYPLGLSYIDDPKAGHSLLYIGILN 213
Db      246 AVIKOSKCEQSTNSYPLVFPFSLLLIGIRVPMISLVPLDDNIKKKSLPAYGALS 305
Qy      214 AINMIGPIIGFTGSLFSKMYVDIGYVDSLIRITPTDSRWVGMWMLNLFVGLFSLSS 273
Db      306 SIRVLGPIGMYMGSCKNFYITLNP-----GLTPADPTWICAMMGFLFGLSIALFPS 361
Qy      274 IPEFLLPQTPNKRQKRSKLSLHLETNDE-EKQDTANLTNQGKNTKNVTVGFPGSPKSI 332
Db      362 TMLFFPQ-----GKEGDS-----AVQLRDVHKERKKKVDDEDRITMMLKDFAKSCKV 412
Qy      333 LTNPLVYMWVLLTLQSSYIAGFTYVFKVVEQOYQOPSSKANILLGVITPIFASGMFL 392
Db      413 LSTRKIMGSVLAGVCVLAFFKGIYVLPYLEHNFPIPLYLVRVYAMFGVGFGLGVAT 472
Qy      393 GGYIIRKFKLNTVGIKAFSCFTAVMSLSFYLLYFFILCENKSVAGLTMTYDGNPNVTSR 452
Db      473 GGYVTKKLINGRAAMFVMLMSTLANVCLXSGKIFIGCS-----IVNSIGNNRQTNVN 527
Qy      453 DVPLSYNSDCNDESOEWBVCNNGITYISPCLACKSS--GSKKPIVFNVCSCLEVT 510
Db      528 FT--RECNSQSCENARLYPVCQDTGFAYFSPCHACREAMQYGSPPVLDFTSCQCAP-G 584
Qy      511 GLONRYSAAHGCPCRDDAC-TRKFFVFAVIAQVNLFFSALGTSHWMLIVKIYQBELKS 569
Db      585 GVSKKPE-----C-ENKCTSSVIFLTV-LPGSFVAGLVVPMALMLRSVPETHS 635
Qy      570 LALGFHSMVIRALGGLIAPYFGLIDTTCIKWSTNNCGTRGSCRTYNSPFSRVYLGSL 629
Db      636 LSLGLQMAVSLPGLTIPSLIMGLVIDAOLVMDKACNGARSGCSLYH-----P 684
Qy      630 SMIRVSLVLYITL-----IYANKKKYQEKDINASNGSYMDS 667
Db      685 DRLRVMMHLLVYVIRFVALITDYYVWKH-----AKGLNIMDR 721

```

## RESULT 9

T26595

hypothetical protein Y32F6B.1 - Caenorhabditis elegans

C/Species: Caenorhabditis elegans

C/Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004

C/Accession: T26595

R/Barlow, K.

submitted to the EMBL Data Library, January 1998

A/Reference number: Z20242

A/Accession: T26595

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-756 &lt;WIL&gt;

A/Cross-references: UNIPROT:O62421; UNIPARC:UPI000007B1F3; EMBL:AL021475; PIDN

C/Genetics:

A/Gene: CESP:Y32F6B.1

A/Map position: 5

A/Introns: 27/2; 148/3; 268/1; 297/1; 349/2; 474/1; 530/2; 585/1; 680/2

Query Match 17.5%; Score 630; DB 2; Length 758;

Best Local Similarity 24.6%; Pred. No. 7.1e-37;

Matches 191; Conservative 132; Mismatches 291; Indels 162; Gaps 24;

```

Qy      30 FLAALSLPIAKTGAIIKSSIIHIERRFEISSILVGFIDSGFEIGNLVIVFVSYFGS 89
Db      7 FIVFGTVFLESIGFPMTSAVVTEKQFLPSRLSGTMVSGADPAIYIPVIFTSYFGG 66
Qy      90 KLAHRPKLIGIGCFIMIGVLTALPHFPMGYVYSKETNIDS-----ENST 136
Db      67 KGNRRARWIGAGIILIANFMILASSNFLFVPEQANTTHIPALAHQIDRIDVNGDVNST 126
Qy      137 STLSTCLINQIISLRASPEIV-----GKG----- 161
Db      127 EHL---WLEQIHPLIDPKQVVMNVEGDEKSDLLQRYVEYCHVYKGSBIKQLKTYIAEK 183

```

```

Qy 162 -----CLKS-----GSTM1VPMGNLRGI 183
Db 184 FPIITDAKSNVRAVVALPYGFCHSMNFVRAOHYACKCKDQSTLTGPFMMI--FGGLLVLG 241
Qy 184 GETPIVPLGLSYIDDPFAKEGHSILYLINALMIGIIFGTGLSFSKMYVIGVYDLS 243
Db 242 GRMTPSLSGLPLMDNVKKNLPLYSFMPFVKILGLVIGLLVGGQNLKLYDPN----P 297
Qy 244 TIRITPTDSRMVGMNLNPLVSLFSLISSIPFFLPQTP---NKQD-----KERKA 292
Db 298 PGQITPLDPMWIGMNLGFLIFGLLPGSLVLYPFPSDDLDINAPBEPDDQKVPK- 356
Qy 293 SLSLHLETDEKQDTANLTNQGKNTKAVTGFQSKSLITPLVYMFVLLTLQVSSY 352
Db 357 --RLNLVDRIHKDEQGNAL--MPETIMDKINDKATIANLVKRIFGAALFGRIDVLA 413
Qy 353 IGAFVYFKYVEOQYQGPSKANILLGVITPIPFASGMFLGYIILKKFKLNTVGIKAFSC 412
Db 414 KGFVPOAKTLELQFVPOYRIQRYIASGTIGFACGVILGSLMKFLQG---RKAA 470
Qy 413 FTAVMSLSFYLLYF--FILCENKSVAG-LTMVTDGNNPVTSHRDVPLSYCNSDCNDES 468
Db 471 WVAVCSILVALISFANGTVGC--KSVIGQIGDOIKNIGPV-----PDGCRDDCKCEQT 521
Qy 469 QMEPVCGNNGITITISPLACG-----KSSGKRIYFVNGSCLEVTGLQRRNYS 518
Db 522 PLYPVCDVSGSAYSPCHACGAMKXYSIFDNTKATVDTLIFENCQVD---NKVKE 576
Qy 519 AHGECPRDACTRKYFFFAIQLVNLFFSALGTSVMILVRIVOBELKSILAGFHSW 578
Db 577 VSRSYC-RTEECBERRFRFMHQAFGAVFGGLGVPEMLILAVAPREHRSVSLGNGFL 635
Qy 579 IRALGGLIAPYIGALIDTTCIKSTNNCGTSCRTYNTSPSRVYL-LSMFAVSSL 637
Db 636 VSLATLPSPIIWKIILMSCLMK--KSCDSGSGCYD--TDELRYALHYICLAFISL 693
Qy 638 VLIILY-----AMKKYOEKDINASENGVDNDEANLSINKK 678
Db 694 ISDVWVWYMARGLKLIBETPBDDKKEBETKLSRKRESLKDAA--EKVERTAY 747

```

## RESULT 10

T27357

hypothetical protein Y70G10A.3 - Caenorhabditis elegans

C/Species: Caenorhabditis elegans

C/Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004

C/Accession: T27357

R./Lloyd, C.

submitted to the EMBL Data Library, October 1998

A/Reference number: Z20354

A/Accession: T27357

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-690 &lt;M.L&gt;

A/Cross-references: UNIPROT:Q9XWCS; UNIPARC:UP10000076276; EMBL:AL032660; PDB:CAA21751.

A/Experimental source: clone Y70G10A

C/Genetics:

A/Genes: CESP:Y70G10A.3

A/Map position: 3

A/Intons: 61/3; 84/2; 185/1; 250/2; 326/3; 375/1; 398/3; 439/2; 490/3; 628/1; 655/1

Query Match 17.2%; Score 618.5; DB 2; Length 690;

Best Local Similarity 26.8%; Pred. No. 4.2e-36;

Matches 169; Conservative 118; Mismatches 292; Indels 51; Gaps 17;

Qy 28 KMFALALST-SPIAKTIGALIMKSIHIERREISSLVGFDGSEIGNLVIYFVS 86

Db 48 KCLLVVLGICAFIQSPVNAIPVGLSTERFRKMTSTHGISSWDFVLLVPEVCH 107

Qy 87 FGSKLHRLKILIGCFPMIGIGVLTALPHFPMGYRYSKKTNTDSSENSTSLTCLINQ 146

Db 108 WGNNGHKGWIGVGMALGSLICALPHMMVDIY---HPDVNLDLTNQTD-FGQC----- 158

```

Qy 147 ILSLNPASPEIYKGLCKSGSYMIVPMGNLRIGETPIVPLGLSYIDDPFAKEGHS 206
Db 159 ---ANRDEBCAGKPHSSWFNPYFMWFI-LGOTLHGVSITPLPSITTYVDENVSGKASP 214
Qy 207 LYLGINALAMIGIIFGTGLSFSKMYVIGVYDLSRTIRPTDSRMVGMNLNPLVSLG 266
Db 215 VYLAHAVALTSRQPIVGVRAQGLNLIDPDRVE--RLPMERSDDRWMGVGIISS 272
Qy 267 LPSIISIPFF-FLPOTPNKPKERKASLSLHLETNDEKQDTANLTNQGKNTKAVTGF 325
Db 273 ISALMLAFPLAFARLPEAKGRADVNGCHVANDVAKAPRDLMLKLPACVWK----- 327
Qy 326 PQSKSLITNPLVYMFVLLTLQVSSYIGAFVYFVYQYQGPSKANILLGVITPI 385
Db 328 -----ILSNPFVLCIFGIESIILNFPALFMPILLETLLSTNPTLASYSVV--I 378
Qy 386 F--ASGMFLGYIILKKFKLNTVGIKAFSCFTAVMSSLPY--LLYFPILCENKSVAGLTMT 441
Db 379 FAATVWVGITIRQLKIQVGMLKMLIVCHVALLFTGLSH---CPQKEFVGINIG 435
Qy 442 YDGNNPVTSHRDVPLSYCNSDCNDESQMEPVCG--NNGITYISPLACGKSSGKKPIV 500
Db 436 YEDLSIERKNDHSISSTCNADGHC--KMEWNPVCDRVTGHMYYSACHAGCTGTITDSSQ 494
Qy 501 FYNCSGLE-----VYGLQRRNYSAHGECPRDACTRKYFFFAIQLVNLFFSALG 552
Db 495 WSGCGCLTNSSTFNLHIGIKHPDVLNQGCHQD--GTYREYILMTPLIYVVASFAG 552
Qy 553 TSHVMLIVRIVOBELKSILAGFHSWIRALGGLIAPYIGALIDTTCIKSTNNCGTSC 612
Db 553 IPIQIMLAVPEPDQKTLALGVNMFVRLLGPIGLIGIILIDVFLCMBG--ESCGKAT 611
Qy 613 CRTNSTSPSRVYLGLSSMLRVSILVYII 642
Db 612 CLVYDPEFKLSWITIGALIVCSILATII 641

```

## RESULT 11

T21217

hypothetical protein F21G4.1 - Caenorhabditis elegans

C/Species: Caenorhabditis elegans

C/Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004

C/Accession: T21217

R./Mortimore, B.

submitted to the EMBL Data Library, October 1996

A/Reference number: Z19392

A/Accession: T21217

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-674 &lt;M.L&gt;

A/Cross-references: UNIPROT:Q93550; UNIPARC:UP1000007841C; EMBL:Z81016; PDB:CA802665.1;

A/Experimental source: clone F21G4

C/Genetics:

A/Genes: CESP:F21G4.1

A/Map position: X

A/Intons: 58/3; 149/1; 181/3; 213/3; 286/1; 327/2; 457/2; 509/3; 608/1

Query Match 16.9%; Score 609; DB 2; Length 674;

Best Local Similarity 25.9%; Pred. No. 1.9e-35;

Matches 169; Conservative 117; Mismatches 283; Indels 84; Gaps 17;

Qy 12 BAQPSSENKTRKCYCNGLKMFALSSPIAKTIGALIMKSIHIERREISSLVGFDG 71

Db 63 ERTPLBGRKKYKKSINIFVLLMLVLAQGTITLGYV--GMULTTERKRRBFSSKSGMILS 120

Qy 72 SFEIGNLVIVVSYFGSLARPKLIGCFPMIGIGVLTALPHFPMGYRYSKKTNTDS 131

Db 121 LVDIGHTMAILLIGYISHYHLPRITGIGVILSSISMFLALPVLVYGTADYTOBQMOK 180

Qy 132 SENSTYSLTCLINQILSLNPASPEIYKGLCKSGSY--NMIVPMGNMLYIGSTPLV 189

Db 181 KE-AVSEVMSCDTN-----GRREISSQGBDCWRHEHHTNAFTIILAFGLFAGIFAPRN 234

```

Qy 190 PGLGLYIDDPKAGHSLV-LGILNAINMIGPIGTCGLSKMYVDIGYVDLSTIRRT 248
Db 235 TIAVYVYISNVQKESPELGLITSMVAFGALGMLSSILNGVTSIGD--ABDHIG 291
Qy 249 PTDSEWVGMWMLNPLVSGLFSSIISSIPPELPQTPKPKOKERKASLSLHLETNDEKQT 308
Db 292 THDEMIGAMWMLGLVCSGSAVILAVPFPFRITRYKKHNT----- 331
Qy 309 ANLTNQGKNTKNTVGFQSGF---KSILTNPLVYMFVLLTLQVSSYI--GAFYVFKY 362
Db 332 -----FLBEFPMVLKGLITNKYITMVIAMMF--GSYLIGGYQVQLPFP 373
Qy 363 VEOQYQSSKANIILGVYITPIFASGMFLGGYIIKKFPLNTVGAIKESCFPA--VMSLS 420
Db 374 IETQGRASAMADIYSGIISVGAIVSTALGWMILSRF--NIAPRSIIICLGSMVILIV 431
Qy 421 FYLYFPLICEKSVAGLT-ATYDGNPVTSHRDVPLSYCNSDCNDESGQWEPVCGNNGI 479
Db 432 SYIIQMLGCSQPKYEGLYVDYASRMHYHREB-DECLCYCNCETVLKFDGVSYNQ 490
Qy 480 TYISPLCAGCKSSSGNKKRPIVFNCSG---LEVTLQNRNYSAMLGECPRDACTRKFY 535
Db 491 NFYSFCHAGCTEYDIYSN--TWSNCCAGVMVVDKGLVHPDCGI-----PF 534
Qy 536 FFWAIQVNLFFSALGGTSHWMLYKIYQPELKSIALGFSHNVIRALGGIAPYFGALI 595
Db 535 AYLAWMLGLFIGNLFFWYTMIVIRSVDEKVALSLASPTIMLFGIIPAPVYGFPI 594
Qy 596 DTYCKMSTNNCGTSCRTYNSTSPSRVYLGSLMRLVRSIVLYIILYANK 648
Db 595 DLCCILMRQCNERNKCVLYNDMFTKMFHGVNSFPQVPAIIFAGICWMLSK 647

```

## RESULT 12

hypothetical protein DKFZp586i0322.1 - human  
 C1Species: Homo sapiens (man)  
 C1Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
 C1Accession: T17250  
 R1Koehrer, K.; Beyer, A.; Mewes, H.W.; Gaasenhuber, J.; Wiemann, S.  
 submitted to the Protein Sequence Database, September 1999  
 A:Reference number: Z18722  
 A:Accession: T17250  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-482 <KOB>  
 A:Cross-references: UNIPROT:O94956; UNIPARC:UP1000002B324; EMBL:AL117465  
 A:Experimental source: adult uterus, clone DKFZp586i0322  
 A:Genetics:  
 A>Note: DKFZp586i0322.1

```

Query Match 14.6% Score 527; DB 2; Length 482;
Best Local Similarity 24.9%; Pred. No. 8-6e-30;
Matches 148; Conservative 96; Mismatches 189; Indels 162; Gaps 18;

Qy 96 LIGGCTFMGIGVLTALPHFPMGYRYSKENTINSSSENSTLTCLINQILSLNRP 155
Db 1 MIGGAILVVALAGLMTLPHFISEPRY-----DNIS----- 32
Qy 156 EIVGKGLCKESGSYMIYVFMGMRLGIGETPIVPLSGYIDDPKAGHSLYGLILNI 215
Db 33 -----FGIS----- 37
Qy 216 AMIGPIIGTGLSLSKMYVDIGYVDLSTIRTPDSEWVGMWMLNPLVSGLFSSISIP 275
Db 38 -----LTIDPRVWGMWMLGLTAAAGVALLAIP 66
Qy 276 FFWLPQTPNKPKOKERKASLSLHLETNDE-----DQTA-----NLTNQGNKI 318
Db 67 YFFPK--EMPEKKEIOPRRKVLAVTDSPPARKGDSPEKSGPSTYKQDGLVQIAPVL 124
Qy 319 TKNTVGFQSGFSILTNPL-YVMFVLLTLQV---SSYIGAFYVFKYVQYQGPSSKA 374

```

```

Db 125 T--VQPIKVFPRVLLQTRHPIFLVVLISQVCLSSMAAGMATTPLPKLERPSTASTA 182
Qy 375 NILGAVITIPFASGMFLGGYIIKKFPLNTVGAIKESCFPAVMSLSFYLYFPLICEKNS 434
Db 183 NLIIQCLSPSVYIVGIVVGGVLRKHLGQVCGALCLLGMILCLFFSLPLFFICSSHQ 242
Qy 435 VAGLTMTYDGNPVTSHRDVPLS-YCNSDCNDESGQWEPVCG-NGGITYISPLCAGCKS- 491
Db 243 IAGIT-----HQTAAHGLSLSPSCMEACSCPLDGFNPVDCPSRVEYITTECHAGCSSW 296
Qy 492 ---SSGNKKPIVFNCSGLEVTGLQNRNYSAMLGECPRDACTRKFYFWAIQVNLFFS 548
Db 297 VQQALDMSQVRYTNCSYV-VEG-----NPVLASG--DSTSHLVFPLLVLSGSLA 348
Qy 549 ALGTSWMLYKIYQPELKSIALGFSHNVIRALGGIAPYFGALIDTTCIKMSTNNCG 608
Db 349 CLTHPSFMLILRGVKKEDKTLAVGIQFWFLRLIAMSPPVHGSALIDTTCVHML-SCG 407
Qy 609 TRGSRTYNSTSPSRVYLGSLMRLVRSIVLYIILYANKKRYQKRDINASNGS 663
Db 408 BRAVCRYNNDLRLNRFLGLQFFFTGSGVTCFALVALVLRQ--QDKKARTKESRS 460

```

## RESULT 13

hypothetical protein K02G10.5 - Caenorhabditis elegans  
 C1Species: Caenorhabditis elegans  
 C1Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 09-Jul-2004  
 C1Accession: T16538  
 R1Leinbach, D.  
 submitted to the EMBL Data Library, November 1995  
 A:Description: The sequence of C. elegans coemid K02G10.  
 A:Reference number: Z18531  
 A:Accession: T16538  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-655 <LBI>  
 A:Cross-references: UNIPROT:Q21157; UNIPARC:UP1000007ABOE; EMBL:U40415; NID:gl065481; PI  
 A:Experimental source: strain Bristol N2; clone K02G10  
 A:Genetics:  
 A:Gene: CESP:K02G10.5  
 A:Map position: X  
 A:introns: 23/2; 113/1; 174/2; 215/1; 248/2; 290/1; 322/3; 371/1; 435/2; 479/3; 60

```

Query Match 13.8% Score 495.5; DB 2; Length 655;
Best Local Similarity 23.4%; Pred. No. 2.1e-27;
Matches 163; Conservative 133; Mismatches 257; Indels 143; Gaps 28;

Qy 27 LKMFPLALS-LSFAKTIGAIT---MKSSIIHERPRISSSLVGFIDSGPIGNLVIY 82
Db 45 LSNFIPMSIISANMLOGALVNGLVSVSISIERFKLTISQGIPTATIDVPTVMLI 104
Qy 83 FVSYGSKLHAPKILGICFCFMIGIGVLTALPHFPMGY--RYSKENTIDS-----SE 133
Db 105 PLALYATATNNKVCIGLGMIVIGISILVITPEYRAGSYGVEKDVCGVGPDKVSE 164
Qy 134 NSTSTLSTCLINQILSLNRPASBEIYVGGCKLCKESGYMIYVFMGMRLGIGETPIVPLG 193
Db 165 GSSDTPFS-----AHREL-----LTLILSGAFVIGASPLFTYGI 198
Qy 194 SYIDDPKAGHSLYGLILNAINMIGPIIGTGLSLSKMYVDIGYVDLSTIRTP----- 249
Db 199 TCLDEFSHKRTGRNALYMLASTYVGPALAPVGCSPMLRLMD-----WRTSPALMG 250
Qy 250 ---TDSRWGAMWMLNPLVSGLFSSIISSIPPELPQTPNKPKOKERKASLSLHLETNDEK 305
Db 251 IDNSADPRFPGIMWVGIVCGFVALFTAPLIMPKR-LKQTVYKA-----NDVH 300
Qy 306 DOTANLTNQGKNTKNTVGFQSGFSILTNPLVYMFVLLTLQVSSYIGAFYVFKYVQ 365
Db 301 RTDAST---DKDFSHKYEFFKIIFMLPKNTKCMCVILMOTIAMLNGYITTFPKLLET 357

```



```

QY 366 QYGPSSKANILLGVTTIPFASGMFLGYIIKKK-----LNTVGIKAFSCPTAV 416
DB 358 LUGFSSGNLSLITGVGVVGVIVASIIIGRISLGFENRKPSPMYFIVIGMAAGSSCL 417
QY 417 MSLSPFLVFLICENKSVAGLMTYDGNPNVTSHPDVP-----LSYCNDCNCDESDQMEP 472
DB 418 L-----IRCEGLNAYGV-----NVPSS---DLRYGALERTSENCHCD-SFPNP 457
QY 473 VCGNNG-IITYISPCLAGCKSSSGNK-KPIVFNCSGLEVTGLQNNRYSAHLBCEPDDAC 530
DB 458 VCSBDSKILFLSPCHAGCADMPKIKFGASNMWTCG-----SKNTIVKGYC--DASC 508
QY 531 TRKPFVFAIVQLNLFPSALGSGSHMLVTKVQPELKLALG-----PHSNV 578
DB 509 QKQIVQFI-----IMFILS-----FCIFTAPVLQSSSLRVNKHKRDHPTCGMLW 556
QY 579 IRALGGILAPIYFAGLIDPTCIKSTNNGCRGSCRTYSTSPSRVYLGLSSMLRVSSIV 638
DB 557 MRLGSIRPAIVFGYIIDVNCMYQ-KDCVSG-KQGFYASNLGMAFPTIAYVKTGCI 614
QY 639 IYIIILYAMKKYQEKD-INASENGSMDEANLESL 673
DB 615 LILFLAAYC---YQESDKSNGKESCRTLETISESV 646

RESULT 14
B81254
probable transmembrane transport protein Cj1588c [imported] - Campylobacter jejuni (stra
C:Species: Campylobacter jejuni
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #ext_change 09-Jul-2004
C:Accession: B81254
R:Parkhill, J.; Wren, B.W.; Mungall, K.; Kelsey, J.M.; Churcher, C.; Basham, D.; Chilling
C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; Vanlille, A.; Whitehead, S.; Barrell
Nature 403, 665-668, 2000
A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hyf
A:Reference number: A81250; MUID:20150912; PMID:10688204
A:Accession: B81254
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-431 <PAR>
A:Cross-references: UNIPROT:Q9PM87; UNIPARC:UPI00000C1F8A; GB:AL139079; GB:AL111168; NID
A:Experimental source: serotype O2, strain NCTC 11168
A:Genetics:
A:Gene: Cj1588c
C:Superfamily: citrate utilization determinant

Query Match 4.9%; Score 178; DB 2; Length 431;
Best Local Similarity 22.1%; Pred. No. 4.4e-05;
Matches 88; Conservative 67; Mismatches 115; Indels 115; Gaps 18;

QY 71 GSFEIGNL---LVIVFVSIFGSKLRRPKLIGICFIMIGCVLTALPHFPMGYRYSKET 127
DB 58 GAFAAGYLARPLGGIVMAHFQDKFGRKN-----FWLSI--LLMVLPTVLAF----- 103
QY 128 NIDSESNSTSLTCLINQILSLNRPSPFIVGKGLKSGSYMMIVVFMGNMLRGISTP 187
DB 104 -----IPGETLGLAPVLL-----ILRIFGIAIG----- 131
QY 188 IVPILGLSYIDDPKAGHSLVYLGIINAMIGPIIG---FTLGLSPSKMYVDIGYVLSLT 244
DB 132 ELPGAMVFAVREYQEKOKAFPLSCINSAVALGILGSIYFLIINAFSIEIRIAY----- 186
QY 245 IRTPTDSRWGAMNLPLVSGFLISSIPFPPLPOTPNKPQKERRKASLSLHVLSTND 304
DB 187 -----AMRIAFVGGIFGIISYLRFRLOETVPFKOMKSSLSFPL----- 229
QY 305 KQOTANLTQGNKITKNTVGFPOSFKS--ILTNPLVMFVL-----TLQVSSYIGAF 357
DB 230 KD-----LPEKDIYVNL---FSSMMMTVLTGCVIVLVLMKPMPSILNLSG----- 275
QY 358 YVFKYVQOYQGPSSKANILLGVTTIPFASGMFLGYIIKKPKLNTVGIKAFSCPTAV 417
DB 276 -----VEGSLQ-----ILGILGIAL--GGAFM-GYLVDFK-----GLFKICIFPSLT 315

```

```

QY 418 SLSPFLVFLICENKSVAGLMTY 442
DB 316 FVFFSFLPYALYELKNLVLVCILY 340

RESULT 15
B84616
hypothetical protein Ab2g22730 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #ext_change 09-Jul-2004
C:Accession: B84616
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, U.A.; Shen, M.; Vanhaken, S.E.; Umayam, L.; Tallon, L.
eaus, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487; PMID:10617197
A:Accession: B84616
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-507 <STO>
A:Cross-references: UNIPROT:Q9ZQ41; UNIPARC:UPI00000A2D4A; GB:AB002093; NID:G4314370; PI
A:Genetics:
A:Gene: Ab2g22730
A:Map position: 2

Query Match 4.9%; Score 178; DB 2; Length 507;
Best Local Similarity 19.7%; Pred. No. 5.3e-05;
Matches 98; Conservative 68; Mismatches 170; Indels 162; Gaps 19;

QY 6 HAKTRAEQPSNKKTRYNGL-KMFLALSLFIKTLGALIMKSIHIERREISS 64
DB 63 YMDRAIASNVGVSSTRSCNDGKCTLAT-----GIQGHFNLNF 102
QY 65 LVGFIDSGFEIGNLVI-VFVSYFASKLRRPKLIGICFIMIGCVLTALPHFPMGYRYS 123
DB 103 EDGVLSSTFVGLLALSPFASL-----AKRLIGVLVMTI-AVL----- 142
QY 124 SKETNIDSESNSTSLTCLINQILSLNRPSPFIVGKGLKSGSYMMIVVFMGNMLRG 183
DB 143 -----GC-----GSSFAFWFIVLCMPFVGV 162
QY 184 GRTPIVPLGLSYIDDPKAGHSLVYLGIINAMIGPIIGFTLGLSFLSKMYVDIGVDS 243
DB 163 GASFTSLAAPPIDNAPQEQKAMWGLFYMCIPSGVALGYVGGVGNF----- 213
QY 244 TIRITPTDSRWGAMNLPLVSGFLISSIPFPPLPOTPNKPQKERRKASLSLHVLSTND 303
DB 214 -----SMRYAFWGEAVLMAFPVVG---FLMKPLQLKGSRTLKNNRLQV-DNEI 259
QY 304 EKDQ-----TANLTQGNKITKNTVGFPOSFKSIILTNPLVMFVLTLQVSSYIGAF 358
DB 260 EHDQFEVSIETSKSSYANAVFSPFGAFDMKVLTKKQFVVNV-LGYVSNFVIGAVSY 318
QY 359 -----VFKYVQOYQGPSSKANILLGVTTIPFASGMFLGYIIKK-----FVL 402
DB 319 WCPKAGNVIK-----MKRADMIRGAVTIIIGIYVTLSSGFTLDRVTATINAPFL 369
QY 403 --NTVGIKAFSCPTAVMSLSFY-----LLYFPLICENKSVAGLMTY 441
DB 370 LSGATFLGAVFCFTTAPFLKSLYGFIALPALGELVFAQTQAPVNAVYCLHCVKSLRPLSMA 429
QY 442 YDGNPNVTSHPDVP-----NVPSS---DLRYGALERTSENCHCD-SFPNP 457
DB 430 I-----STVALHIFGVDVPS 444

Search completed: March 14, 2006, 10:52:50
Job time : 49 secs

```